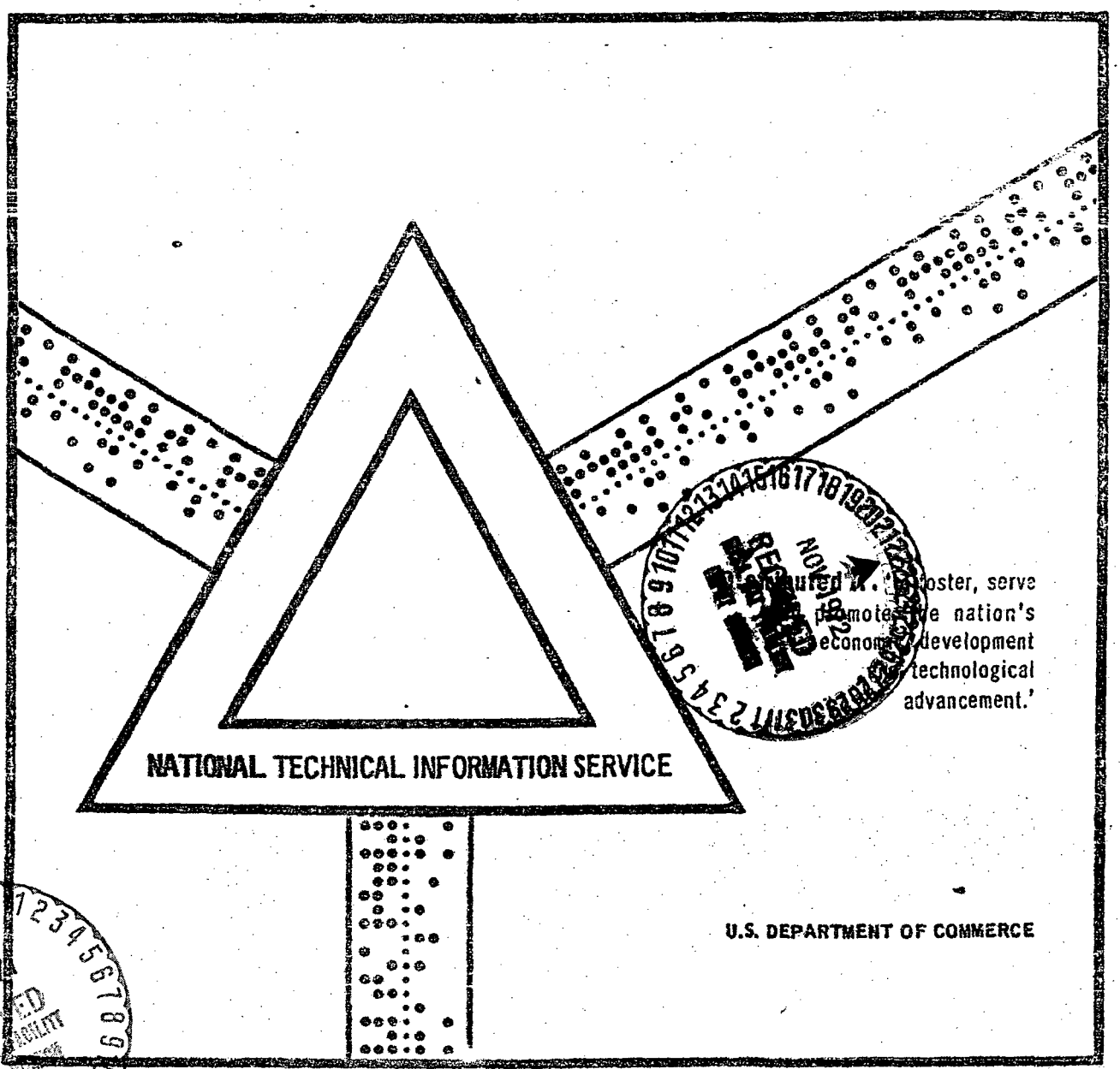
REPORT OF THE AD HOC GROUP ON DATA CENTERS

James I. Vette, et al

National Aeronautics and Space Administration
Greenbelt, Maryland

September 1969



REPORT OF THE AD HOC GROUP ON DATA CENTERS

Committee on Scientific and Technical Information
Federal Council for Science and Technology

September 1969

REPORT OF THE AD HOC GROUP ON DATA CENTERS

Prepared By:

Dr. James I. Vette (Chairman), Director
National Space Science Data Center
GSFC

Dr. Woodrow C. Jacobs, Director
Environmental Data Services
ESSA

Dr. Thomas S. Austin, Director
National Oceanographic Data Center
U.S. Naval Oceanographic Office

Dr. Donald W. Pritchard, Director
Chesapeake Bay Institute
The Johns Hopkins University

Dr. George H. Ludwig, Chief
Information Processing Division
Tracking and Data Systems Directorate
GSFC

Mr. James A. Fava (Executive Secretary), Deputy Director
National Space Science Data Center
GSFC

September 1969

FOREWORD

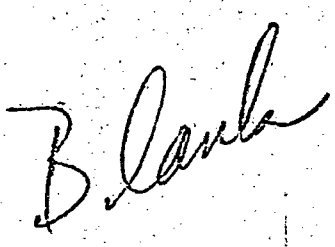
Our national evolution toward a post-industrial society has many specific features which must prescribe changes in the way we live. One of the more significant of these is the recognition that we now live in an information-rich world. Twenty-five years ago, information was a scarce commodity—now we are flooded with more information than we can use or even handle, both as private citizens and especially as members of the scientific community. Indeed, for the typical problem-oriented scientist or engineer, the volume of relevant or important information is so great that it constitutes a barrier to its own use.

This report considers the problems of one kind of information resource—the data center which receives a large mass of data points. This report is especially relevant because the data centers studied, as representative of this class, are those concerned with environmental sciences. Thus, this report anticipates and seeks to ameliorate some concerns which will be vital to our country throughout the decade of the 70's, and beyond.

It is perhaps appropriate to observe that data centers of the type discussed can perform two functions. One is immediate and mission-directed; here the center does the things which its sponsors require as essential for their mission accomplishment; the other function is secondary, and seeks to make further dissemination and hopefully utilization (for a wider user audience) of the data which were collected to perform the first function. This report addresses itself solely to the secondary, broad-based utilization of data, since the problems in this area are by far the greater. A considerable amount of study has been given this report since its preparation, largely because of the increasing attention being devoted to environmental pollution, earth resource satellites, and marine resources. It is now being released to provide guidance and direction to those seeking to solve complex data problems in these fields.

Andrew A. Aines

Andrew A. Aines
Chairman, COSATI

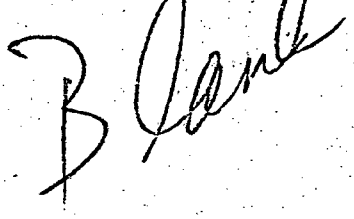


CONTENTS

		Page
I.	INTRODUCTION	1
II.	GENERAL DESCRIPTION OF DATA AND INFORMATION FLOW	2
	A. Information Flow	3
	B. Characteristics of a Data Center	5
	1. Acquisition	6
	2. Storage and Retrieval	6
	3. Analysis	7
	4. User Services and Products	8
	C. General Remarks on Existing Environmental Data Centers	8
	D. Relations Between Existing Data Centers	9
	E. Requirement for Additional Data Centers	10
	F. Position of a Data Center in the General Information System	10
III.	PROBLEMS CONFRONTING DATA CENTERS	10
	A. Resources	10
	B. Availability of Data to Users	11
	C. Technical Standards	12
IV.	RECOMMENDATIONS	12

ILLUSTRATION

Figure		Page
1	Data Flow	4



REPORT OF THE AD HOC GROUP ON DATA CENTERS

I. INTRODUCTION

Man has been communicating and preserving the results of his labors and the increase in his knowledge through the written word or symbol for many centuries. In general, this process has condensed the results of many hours or years of effort into a form which others can understand and use with a considerably reduced expenditure of human resources. However, with the increase of population, wealth, and the facility of communication, the generation of new facts and knowledge resulted in such a stockpile of information that methods had to be devised to store and retrieve this product so that individuals could be provided only with the relevant information they needed for specific purposes. Volumes have been written during the past 10 years concerning the information explosion and the ever-increasing number of publications. Numerous systems have been established to acquire, abstract, keyword, announce, store, retrieve, and furnish copies of publications to requesters. Numerous groups have studied, identified, and solved many of the problems associated with these areas.

With the advance of modern technology in computers and data storage devices, interest has turned from the problem of handling the results of an experiment or operation in written form to the preservation and dissemination of the data produced in the conduct of that endeavor. Since this latter task increases the information content to be handled by one hundred to one thousand fold over the written report, there should be concern about the problems that are encountered and how these should be solved on a national and international level. Unfortunately during the past decade when this type of data has been accumulating at an unprecedented rate, insufficient attention has been focused on the problems associated with overall data management. Data centers have been established to handle this task in certain fields. Such centers differ from the large documentation centers because they deal with huge masses of quantitative measurements which, in analog or digital form, may exist on microfilm, hard copy, computer printouts and plots, magnetic tape, etc. In many instances the initial use of the measurements has already been exploited; subsequent use often requires professional expertise in both data processing and the associated scientific discipline.

A comparison between present data collection, reduction, and evaluation technology and that of 10 years ago illustrates how the problems of the data centers have been magnified during this decade. Before the use of computers, it was customary to display the raw data by manually recording individual measurements or by using chart recorders. The data generators and primary data users, such as scientists and engineers conducting research or an operational

mission, would then work with those individual measurements using desk calculators, slide rules and pencil and paper to reach their conclusions or present their data.

With the advent of computer technology, these users have been relieved of the problem of manipulating the individual measurements and are able to specify the repetitive operations to be performed by these devices. Therefore, they are able to work effectively with data bases which are expanded by factors of thousands to millions.

At the same time, the number of these active data generators and primary users has increased tremendously thereby making the direct exchange of data between individual researchers more difficult. These factors, plus the present tendency for research efforts to cut across disciplinary boundaries, result in the present increased need for data centers to have extensive capabilities and full-time professional staffs.

An ad hoc group was established by the Committee on Scientific and Technical Information (COSATI) to examine the problems associated with centers charged with the responsibility of handling large data bases and to make recommendations on those problems which require attention by executive groups. In order to get at the heart of this quickly without making an exhaustive study of all types of data centers, it was decided to study data centers associated with the environmental sciences. The volume of data generated in these fields is presently larger than in most others, and the diversity of their user community is quite extensive. Consequently it was felt that the results of this study should have a broad applicability to other fields, e.g., medicine, social science, education, etc.

This report is the result of a short-term study conducted by the ad hoc group composed of individuals with backgrounds in the various fields of environmental sciences. Since it was established at the outset that a large data center is an important element of any information system serving a definable segment of the environmental sciences, the report contains a description of a generalized data center, a discussion of the broad functions which this center should perform, and the relationship of this activity to the general flow of information throughout the professional and user communities associated with the particular discipline. With this as a background a number of common problems are identified in Section III. The group's recommendations are given in Section IV.

II. GENERAL DESCRIPTION OF DATA AND INFORMATION FLOW

Environmental science data are produced from quantitative measurements of phenomena taking place within the environment of the earth and interplanetary

space. The bulk of these data originate within the disciplines of geophysics—aeronomy, meteorology, hydrology, oceanography, seismology, geomagnetism, geodesy, and the extension of these into space, i.e., space and planetary sciences.

Environmental data may be collected for a number of reasons. The motivation may be one of basic research in which an attempt is made to find out what is there, how it varies with time and space, as well as to understand its properties in terms of fundamental processes and principles. On the other hand, there may be an operational mission which must be supported, or the data may be collected for an economic need. Regardless of the initial motivation, much of the data, either in the fundamental or in a converted form, may be very useful to others—and for entirely different reasons. Since these data are expensive and time consuming to obtain, their preservation for additional use is important. In order for this preservation to be justified economically, the costs for such activities must be a reasonably small fraction of the original costs for obtaining, processing, and analyzing the data.

A. Information Flow

For purposes of this report, a single environmental data measurement performed at a given location and time becomes a data or station point. A data point can be considered as a unit of fundamental information obtained from a sensor. The space-time point and the appropriate characteristics of the measuring devices constitute associated information necessary to use this physical, chemical, or biological measurement.

Once data are obtained, say at a weather station, from an oceanographic survey ship, satellite, etc., some initial preparation may have to be accomplished to make the data useful. They may pass through an acquisition station and be relayed over a communication link to a processing facility. There mechanical, electrical, computational, or other techniques may be applied in order to change the data from one form to another, e.g., analog to digital. The data could then flow to an experimenter or, for use in a real-time mode, to an operational unit. In both instances the data may be processed and thus reduced into a useful, ordered, or simplified form for operational purposes or for scientific analysis. An idealized picture of data flow is shown in Figure 1.

The type and amount of data that flow into a center depend upon the mission and nature of the center. The actual time involved in the flow from source to the center may range from hours in the case of synoptic weather data to years in the case of oceanographic or satellite data, where individual scientists are responsible for the general conduct of the experiment and the subsequent primary analysis of the data. In any event good and valid data with the necessary

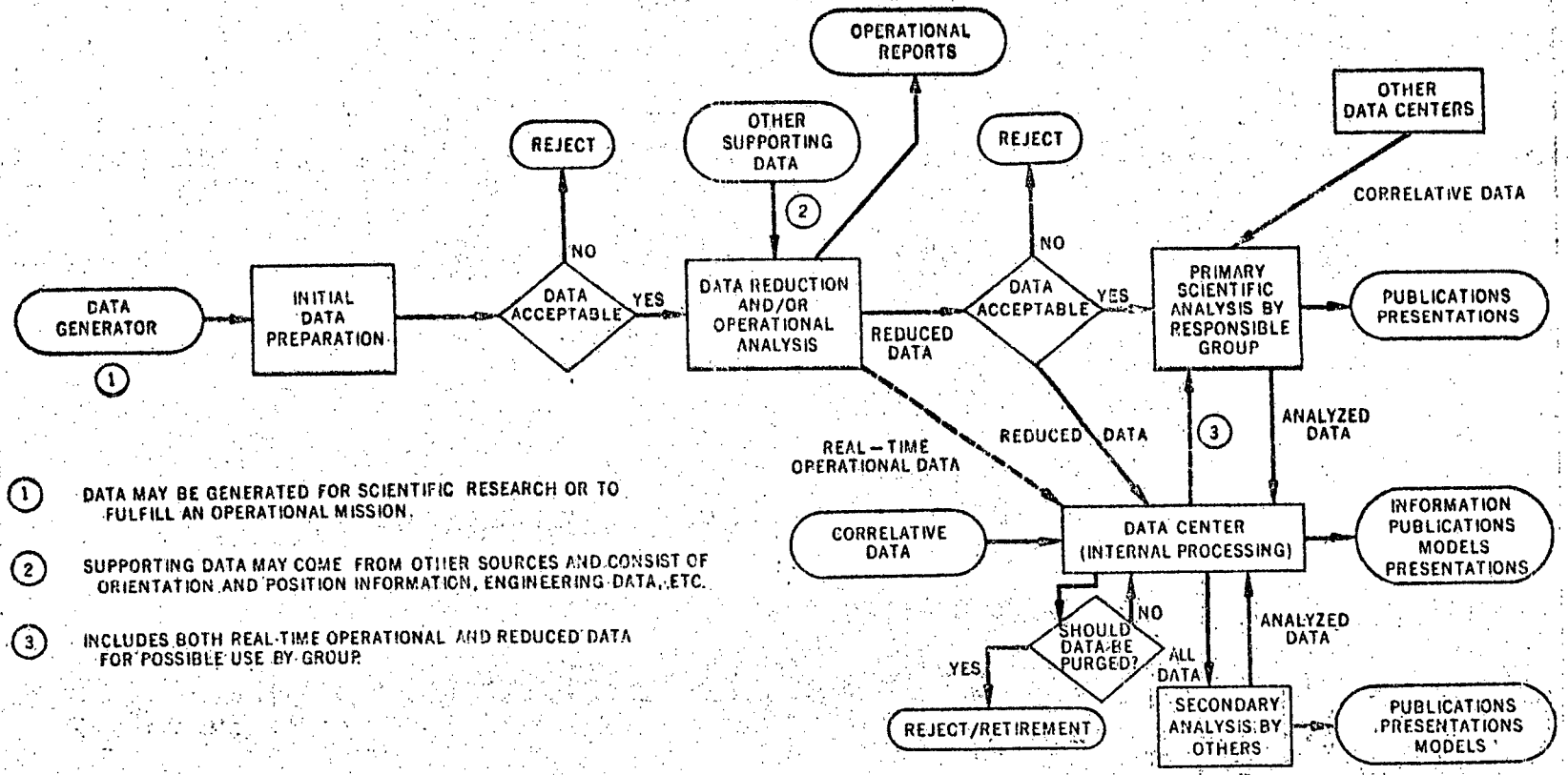


Figure 1. Data Flow

documentation to adequately describe the experiment and the characteristics of the measuring sensors should reach the appropriate center. It is not necessary or feasible in some instances for the center to acquire all useful data. By maintaining a directory of specialized data bases the data center may call upon, or refer the user to, these peripheral data collections.

B. Characteristics of a Data Center

Although individual data centers have their unique characteristics, they also have features which are common and necessary to discharge the functions that will be described more fully below. A data center, although discipline oriented (e.g., to meteorology, space, oceanography), is responsible for the archiving and subsequent use of the data obtained from a particular segment of the scientific community or a data generation activity.

In order to perform this mission, the center must naturally acquire appropriate data and the necessary correlative information and documentation. If the data cannot be handled by a diversified spectrum of users with a minimum of effort, they should remain with the original investigators and be noted as available. The data center should have at least the following capabilities:

- An information system about both the data in the center and the availability of the specialized data collections that exist in other locations. Microfilming, digitizing, and computing equipment with enough flexibility to be able to accept data in most any form or format and be able to provide the data in a variety of ways so that it is readily usable by a diversified user community.

- A specialized technical library and automated document retrieval system. Since the generalized documentation systems cannot be used efficiently as specialized systems, it is necessary to develop a descriptor hierarchy suitable for each discipline and oriented toward the professional user.

- A professional staff in the scientific disciplines that carries on analysis and synthesis of the data. The main direction of this effort should be such that the end products are (a) new and useful forms, (b) summaries and compilations, (c) model environments, and (d) state-of-the-art reviews. The center is then capable of serving as an information center in the appropriate area of science.

- A professional staff in the computer and information sciences that develops information systems, analysis routines, storage and retrieval techniques based on latest capabilities in computers, data storage devices, communication links, and interactive input/output devices.

Four of the more important functions which require detailed discussion are: (1) acquisition, (2) storage and retrieval, (3) analysis, (4) user services and products. A data center must concern itself with the quality control of its data while accomplishing all of its functions. In addition, it has to keep abreast and even involve itself in the technical development of data processing techniques and equipment.

1. Acquisition—To be successful a data center must have a very active acquisition effort. Those responsible for acquisition must be professionals, technically competent in their disciplines. During the early planning phases of any large-scale, data-gathering programs—whether for research, survey, or operational purposes—the acquisition specialists of the appropriate center(s) should be involved in order that processing techniques will be used to optimize use of the data both for the goals of the program and for the input/output cycle of the center. In addition, the collection of the necessary correlative data can be anticipated at this time. Individuals involved with smaller scale research efforts should be advised by the center as to the best means to preserve the data for use by others. A flexible input/output system of the data center is of great advantage in communicating this data to a wide variety of users.

Once a data gathering program is approved, the acquisition staff must start working with the generators during the time that data reduction plans are being formulated. It is at this time that the function of the center and the problems associated with archiving the data must be clearly understood by the generators. While working with the generators, the center representatives must maintain a flexible but persistent schedule. This schedule should allow for the rejection of data of questionable quality and of data with inadequate documentation and allow for slippages of program schedules. The data to be submitted to a center should be in a form which requires the least expenditures of resources—money, manpower, computer time, etc., considering both the data generator and the data center. Normally this form of data will be a natural product of the data processing and only needs to be preserved at the proper point in the cycle.

2. Storage and Retrieval—After the receipt of the data and documentation as discussed in the previous paragraphs a data center must perform a number of operations on them in order for these to be stored in such a manner that they can be readily retrieved and used by others. The data and documentation must be appropriately identified and properly routed within a center. It must be descriptively cataloged and then indexed according to a system appropriate for that data center. The search strategies employed for data retrieval must be able to retrieve particular data sets or subsets by translating a user's request to a form which allows for specialized selection of the desired data portions with all attendant information needed for its use. While the actual strategies may vary from

center to center, consideration should be given to retrieving data by vehicle (spacecraft, aircraft, ship, etc.), time, geographical position, sensor, experiment, experimenter, operational system, etc.

A data center may collect data which has been recorded on (a) microfilm, (b) digital magnetic tapes, (c) photographic positives and negatives, (d) graphs and roll charts, (e) microfiche, (f) computer generated plots, (g) printed material. Since it is necessary to have special-purpose equipment to handle analog tape data, a center should not normally be expected to accept such data. In order to conserve resources and to store the data in the most appropriate form, it may be necessary to convert the data from one form into another form.

During these processes the data must receive a quality check to determine if the actual data content and the documentation on the data are accurate and correct. It is at this stage that any questions concerning quality should be clarified. There is no point for a data center to expend time and effort archiving data of a questionable quality. Eventually, the data and the necessary documentation are stored and cataloged so that they can be recalled for future use.

3. Analysis—Environmental data centers—while maintaining and improving the acquisition, storage, and retrieval functions—should actively develop a strong capability for analysis to meet the user needs for various data products. The end products of such analysis (and synthesis) should be new and useful products, compilations, or models which are desired by the user community. Only in this way will centers be able to attract professionals of sufficient competence in the various disciplines to guarantee the proper data inputs and internal data management. The creation and documentation of a particular model of some environmental parameters can be considered as a state-of-the-art survey in a scientific field as well as a useful new output. Such a model, in lieu of a well-developed theory, may make certain classes or groups of data redundant, resulting in a possible compaction of the data, or may serve to identify certain data as no longer useful. Thus these data subsets could be retired from the active data base or purged completely. It is clear that any high-volume data center must, of economic necessity, establish a data retirement or purging system. It would be neither wise nor economical to acquire and archive forever all types of data. However, decisions involving purging or retirement should normally be left to the judgment of professionals and not be made by an arbitrary agency policy or procedure.

It is only logical that once a data center develops a strong capability for analysis, several information analysis centers will evolve within the data center. It must be realized that both the analysis and information-type functions require a number of years to develop. The data center must reach a certain minimum size, both as to resources and the types and amounts of data, before it can really

become effective in these fields. This minimum size will depend upon both the discipline(s) associated with the center and the segment of the scientific community to which the center is responsive.

4. User Services and Products—There is really no purpose in having a data center if the center cannot provide a wide variety of services and products to users. Both users and data center managers, however, must realize that a center will never have sufficient resources to satisfy all user demands for service. A data center may be able to recover the cost of its services; however, the input and internal development costs could not be recovered.

Such services should include, but are certainly not limited to, the following:

a. Disseminating catalogs and data center publications

b. Retrieving, reformatting, and furnishing data

c. Furnishing necessary space and use of facilities for visiting scientists

d. Furnishing special bibliographies

e. Preparing and publishing models

f. Evaluating and analyzing data to meet individual requests

g. Summarizing and preparing graphic displays

h. Providing directories and referral services

i. Consulting, reducing, and processing data

In many instances the major secondary users do not require the data per se, but require products that are derived from extracting, compiling, evaluating, reformatting, and synthesizing the data. Such products may be charts, atlases, models, statistical studies of properties and phenomena, handbooks, etc. The users of such products in all probability may not be the scientists intimately involved in the particular discipline. More commonly they include such groups as (a) scientists in related disciplines, (b) engineers and designers, (c) planners, (d) management, (e) operational activities, (f) educational activities, (g) recreational activities, (h) commercial activities, and (i) general public.

C. General Remarks on Existing Environmental Data Centers

At the present time there are a number of centers in the United States concerned with environmental data; however, it does not seem essential to list them

explicitly in this report. It should be pointed out that the most effective of these centers are mainly discipline oriented and not agency oriented. The group felt very strongly that this pattern should be encouraged and further developed. It is also felt that the responsibility for the archiving of data from a particular program, regardless of the agency supporting the program, should go to the appropriate disciplinary data center. Thus interagency cooperation is essential to an efficient overall Data Management System.

During the initial phases of planning for a particular environmental data-gathering program, a determination should be made of the advisability of preserving the data for secondary use. If this decision is positive, the proper data from each experiment and/or survey operation should be acquired by the selected data center. With this advanced determination the center would then have sufficient time to prepare for the receipt of the new data. In addition this procedure would readily identify those cases where an appropriate center for a particular class of data does not exist. At that time a decision could be made either to initiate a new facility or to expand the mission of an existing center. Since there are many government agencies involved with the funding of research and operational programs as well as carrying out some of these programs, a coordination group or executive body would probably have to play a role in these determinations.

D. Relations Between Existing Data Centers

Although the group felt that an adequately supported, centralized data center is a necessary and important element of any overall information system serving a particular segment of the environmental sciences, there currently are no requirements for a monolithic data center or for high-speed data links among all existing data centers. There is, however, a genuine need for close coordination and cooperation among existing environmental data centers. A focal body should be developed to expand the desired level of coordination and to facilitate the development and spread of technological advances in storage, manipulation, and retrieval. This body would be mainly composed of the directors or appropriate representatives of the existing data centers and members of the user community. It could also coordinate activities between overlapping scientific fields and help identify disciplines in which expanded or new data centers are needed.

Each of the existing centers should be aware of the holdings and services of the others so that requests may be funneled to the correct center for action. It is quite possible, with advances in high-density storage media, that high-speed links to data on line will soon become economically feasible.

E. Requirement for Additional Data Centers

While the group did not make a searching and thorough analysis, it did conclude that there are a number of types of environmental data for which there are no national centers at the present. Those identified were:

1. Ground-based visual, radio, and radar observations of the planets

2. Earth resources data originating from satellites

3. Solid earth geophysics data

F. Position of a Data Center in the General Information System

It should be emphasized that a data center does not replace any element in a well-organized information system serving a particular scientific discipline. The center merely represents a new addition in the overall system and is essential in those fields where vast amounts of data are generated at considerable expense which have wide use outside the specialized scientific or operational activity which generates such data. The professional societies, meetings, and publications in journals (both scientific and trade) still provide the primary communication of information within the discipline and its peripheral areas. The mission-oriented and cross-disciplinary information analysis centers are not replaced by the activities of the large data center. On the other hand, by virtue of its essential analysis functions, the data center described here contains within its structure a number of information analysis centers focused mainly in the scientific disciplines.

III. PROBLEMS CONFRONTING DATA CENTERS

The group did not go into the internal problems that are unique to a data center and may be even unique to a particular set of data, operational system, or experiment. It is the judgment of the group that such problems are peculiar to a particular center and can be solved given time and resources. Therefore, only general problems which cannot be totally solved by the data centers themselves are discussed here. These have been generalized as (a) resources, (b) availability of data to users, and (c) technical standards. Each is discussed in the following paragraphs.

A. Resources

The group did not assess the capabilities of the existing data centers; however, it is known that each center does not have sufficient resources—money,

manpower, facilities, and equipment—to adequately carry out its assigned responsibilities. Therefore, it is imperative that each center have some options in deferring or accepting past and currently available data. They should have the prerogative of determining what data are important in terms of known or potential user requirements and on what they should expend their limited resources.

The group did not feel that data centers could be totally self-sufficient in the same sense that research and development efforts are not. The agency responsible for the data-gathering program should provide funds for the experimenter(s) and/or operational programs to make the data and documentation available to the center. The agency responsible for the operation of the center should fund for the internal operation and for its portion of the acquisition costs. It would be appropriate for a fraction of the agencies' R&D and/or operational budget which supports the data-gathering programs to be used for supporting data center activities. In order to have a data center with the capabilities described earlier, usually between 1-5% of the total funds expended to generate the data for primary use are required to support the center. If the projected use of the data beyond its primary function is not great enough to warrant this cost, then a data center approach is not practical. The existence of such centers in the environmental sciences demonstrates there are cases where this type of operation has proved effective.

In order to reduce the overall government cost of operating data centers and to prevent abuse of its services each should be able to charge for its outputs and should have the means of using these funds. A uniform user charge policy is highly desirable for federally operated data centers. There are difficulties in getting various agencies to agree on a uniform price for a particular medium, e.g., the price for reproducing a 100-foot reel of 35-mm microfilm, since unit costs vary from center to center. Some central body within the government should resolve this problem by providing a funding mechanism for achieving a balance between cash receipts and costs.

One of the management problems associated with the data center is finding and attracting qualified personnel to the data center field because it is much more glamorous and exciting to be generating the data and performing the primary analysis. Since the professional activities of the center are sufficiently closely related to the scientific disciplines which they are serving, personnel can be recruited from those disciplines, and no special "breed" of professionals is required.

B. Availability of Data to Users

Except for those special centers handling data which are vital to national security or are of distinct commercial value, the interchange of data on an

international level should be encouraged. For government funded data centers, it is clear that any U.S. citizen is allowed to purchase the output, except, of course, for classified data. Futhermore, the ability of data centers to supply their outputs free in limited quantities for educational and scientific uses seems to be desirable and in the best economic interests of the government. World data centers have been established to facilitate the international exchange of environmental data. National data centers should continue to support their operation as an effective means of overcoming bilateral exchange restrictions.

C. Technical Standards

The group endorsed the idea of having technical standards for data handling and manipulation, reproduction, and storage equipment to simplify the problems of using the data base over an extended time period. In addition, such standards would allow a much greater ability to service a wide variety of users without excessive costs. This problem is not unique to a data center but is general throughout the data processing field. The group felt very strongly that a rigid data format standard should not be established. There a reasonable degree of flexibility should be maintained.

IV. RECOMMENDATIONS

In this brief study, emphasis has been placed on identifying problems associated with large-volume data centers which require the attention of executive-level groups. These problems require more detailed examination and a continuing review beyond the capability of an ad hoc group. In order to accomplish these functions, and to keep COSATI and the Office of Science and Technology (OST) informed on a continuing basis, the following recommendations are made:

A. The Office of Science and Technology should be encouraged to establish broad policies, objectives, and procedures to insure that:

1. Large national discipline-oriented data centers be established in the appropriate agency to handle the dissemination of useful data to all secondary users.

2. The agency responsible for a given national data center provide the necessary funding to adequately develop the center so that it can perform services for all agencies and users.

3. Each Federal agency that supports or conducts programs to obtain large amounts of data be made aware of the economic value

attached to the secondary use of this information; accordingly it should devote a portion of its R&D and/or operational funds to allow for the ultimate transfer of this data and supporting information to the appropriate national data center.

4. Each Federal agency should establish procedures whereby representatives of national data centers would work with the program planners and project personnel of the different agencies during the initial phases of a large data-gathering program; the goal should be an efficient transfer of the data appropriate for secondary use and retention of valuable information to the appropriate data center in a timely manner. As projects phase out every effort should be made to acquire and archive appropriate data.

B. That COSATI should establish a Panel on National Data Centers. Initially the Panel could be composed of the directors of the large environmental data centers and selected individuals from the user community. The Panel's responsibilities would include but not be limited to the following items:

1. To keep OST apprised of the major external problems confronting national data centers and the progress of these centers in achieving their role in the overall information system.

2. To encourage the continuation of the emerging pattern of discipline-oriented rather than agency-oriented data center and advise OST on the establishment of new national centers.

3. To support the requirements of the national centers.

4. To coordinate and assist in resolving their common problems.

5. To facilitate the incorporation of the technological advances in computers, high-density storage and retrieval, and communications into the national centers.

6. To report the findings of the Panel to COSATI.

C. COSATI should establish an ad hoc group to explore the possibilities of:

1. A uniform service-charge policy for all government-operated or supported data/document services

2. Standard unit prices

3. Revolving or trust funds for all government-operated or supported data/document services organizations, and

4. Alternative improvements in procedures.

STANDARD TITLE PAGE FOR TECHNICAL REPORTS	1. Report No. COSATI-70-5	2. Govt. Accession No.	3. Recipient's Catalog No.
4. Title and Subtitle Report of The Ad Hoc Group on Data Centers			5. Report Date Sept. 1970 6. Performing Organization Code
7. Author(s) See below			8. Performing Organization Rept. No.
9. Performing Organization Name and Address COSATI Ad Hoc Group on Data Centers			10. Project/Task/Work Unit No. 11. Contract/Grant No.
12. Sponsoring Agency Name and Address Committee on Scientific and Technical Information Executive Office Building Washington, D.C.			13. Type of Report & Period Covered Task Group Study 1969 14. Sponsoring Agency Code

15. Supplementary Notes Prepared by Dr. James I. Vette (GSFC), Dr. Woodrow C. Jacobs (ESSA), Dr. Thomas S. Austin (U.S. Naval Oceanographic Office), Dr. Donald W. Pritchard (The Johns Hopkins Univ.), Dr. George H. Ludwig (GSFC), and Mr. James A. Fava (GSFC).

16. Abstracts This report deals with the problems associated with data centers charged with the responsibility of handling large data bases. Data centers concerned with environmental sciences were studied; however, because of the large volume of data generated in these fields and the diversity of the user community, it was felt that the results of this study could have a broad applicability to other fields such as medicine, social science, education, etc. Regardless of the original motivation for collecting data, much of the data may be very useful to others - and for entirely different reasons. Since these data are expensive and time consuming to obtain, their preservation for additional use is important; the cost of such preservation and use should be a reasonably small fraction (1-5%) of the original costs for obtaining, processing, and analyzing the data. It is recommended that: 1. The Office of Science and Technology be encouraged to establish broad data center policies, objectives, and procedures. 2. COSATI establish a Panel on National Data Centers and an ad hoc group to study a uniform service-charge policy, standard unit prices, and revolving/trust funds.

17. Key Words and Document Analysis. 17a. Descriptors

Data acquisition, data analysis, surveys, data processing, data reduction, data retrieval, data storage, information systems, data transmission, environments, cost analysis, cost centers, cost control, cost effectiveness, cost engineering, cost estimates, national government.

17b. Identifiers/Open-Ended Terms

COSATI, data center, secondary data users, national data centers, revolving-trust fund.

17c. COSATI Field/Group Electronics and Electrical Engineering, Information Theory; Earth Sciences & Oceanography.

18. Distribution Statement	19. Security Class (This Report)	21. No. of Pages
Unlimited	UNCLASSIFIED	17
	20. Security Class (This Page) UNCLASSIFIED	22. Price